



## INFERÊNCIA SOBRE A ESTRUTURA POPULACIONAL DA SOJA COMERCIALIZADA NO BRASIL, USANDO DADOS GENOTÍPICOS, PARA USO NO MAPEAMENTO ASSOCIATIVO DE ALTO TEOR DE PROTEÍNA E MAIOR EFICIÊNCIA NA FIXAÇÃO BIOLÓGICA DE NITROGÊNIO

INFERENCE ABOUT POPULATION STRUCTURE OF SOYBEAN MARKETING IN BRAZIL, USING GENOTYPIC DATA, FOR USE IN ASSOCIATION MAPPING OF HIGH PROTEIN CONTENT AND HIGHER EFFICIENCY IN BIOLOGICAL NITROGEN FIXATION

TORRES, A.R.<sup>1</sup>; GRUNVALD, A.K.<sup>1</sup>; MARTINS, T.B.<sup>2</sup>; SANTOS, M.A.<sup>1</sup>; LEMOS, N.G.<sup>1</sup>; SILVA, L.A.S.<sup>3</sup>; HUNGRIA, M.<sup>4</sup>.

<sup>1</sup> Bolsista Recursos Humanos CNPq, programa RHAE, Embrapa Soja, Londrina, PR. E-mail: [adalgisa@cnpso.embrapa.br](mailto:adalgisa@cnpso.embrapa.br)

<sup>2</sup> Bolsista DTI CNPq, Embrapa Soja, Londrina, PR;

<sup>3</sup> Soytech Seeds, Rio Verde, GO;

<sup>4</sup> Embrapa Soja, Londrina, PR.

### Resumo

Uma dificuldade encontrada nos programas de melhoramento de soja é a de aumentar o teor de proteína sem reduzir a produtividade de grãos. Para contornar essa dificuldade, uma alternativa seria melhorar a eficiência da fixação biológica de nitrogênio (FBN). A associação simbiótica entre a soja e bactérias das espécies *Bradyrhizobium japonicum* e *B. elkanii* tem importância mundial para a cultura. O presente trabalho teve como objetivo avaliar a estrutura populacional das principais cultivares de soja comercializadas no Brasil, usando dados genotípicos, para uso no mapeamento associativo de alto teor de proteína e maior eficiência na FBN. O experimento foi realizado no Laboratório de Biotecnologia do Solo da Embrapa Soja. Um total de 192 cultivares de soja foram genotipadas com 21 marcadores SSR relacionados a QTLs que controlam teor de proteína em soja. A genotipagem foi feita em geis de poliacrilamida a 10% e oito marcadores não ligados foram selecionados para análise de estrutura populacional com o *software* Structure. Foi testada a hipótese de 1 a 10 subpopulações, com mistura e frequências alélicas correlacionadas. O número de alelos na população variou entre 2 e 11, com uma média de 4,3 alelos por locus. Somente dois marcadores não revelaram polimorfismo. Os resultados mostraram que a população está estruturada, com três subpopulações distintas entre as cultivares.

### Introdução

A fixação biológica de nitrogênio (FBN) é considerada uma alternativa para o aumento do teor de proteína no grão de soja. No Brasil, a FBN contribui significativamente para a produção de soja e representa uma economia estimada em US\$ 7 bilhões por safra (HUNGRIA et al., 2007). Tanto o teor de proteína quanto a FBN são características quantitativas e, portanto, possuem um controle genético complexo. Vários estudos de mapeamento de QTLs (*Quantitative Trait Loci*) relacionados com o teor de proteína nos grãos de soja são descritos, contabilizando mais de 100 QTLs (SOYBASE, 2010). Apesar de um grande número de QTLs mapeados, a maioria foi detectada em um número pequeno de populações e em poucos ambientes. Por outro lado, poucos estudos sobre QTLs controlando a FBN foram realizados (TANYA et al., 2005; NICOLÁS et al., 2006; SANTOS et al., 2006), e a maioria dos QTLs encontrados coincide com regiões onde foram descritos QTLs para proteína. Até o momento não foram realizados estudos que abordassem, ao mesmo tempo, o aumento do teor de proteína no grão e a FBN, persistindo a necessidade de aprofundar o entendimento sobre as bases genéticas desses dois caracteres.

O conhecimento da estrutura da população e/ou parentesco entre as cultivares é imprescindível para o estudo de mapeamento associativo, pois minimiza possíveis erros de associação entre os caracteres. Além disso, segundo ROSENBERG et al. (2002), a análise genotípica multilocus permite inferir sobre os ancestrais, sem depender de informação sobre o local de amostragem das cultivares.

## Material e Métodos

O experimento foi realizado no Laboratório de Biotecnologia do Solo da Embrapa Soja, em Londrina, PR. Foram utilizadas 192 cultivares de soja atualmente comercializadas no Brasil e desenvolvidas por instituições públicas e privadas. Para a extração de DNA, as sementes foram semeadas em casa de vegetação. Após 2-3 semanas de crescimento, folhas jovens foram coletadas e armazenadas a  $-80^{\circ}\text{C}$ . O DNA foi extraído de acordo com o protocolo descrito por KEIM et al. (1988). A genotipagem seletiva foi feita inicialmente (SANTOS 2006) para encontrar prováveis associações entre marcador/FBN. Para a avaliação de polimorfismo entre as cultivares foram utilizados 21 SSR (Tabela 1), desenvolvidos por CREGAN et al. (1999) e relacionados a QTLs que controlam teor de proteína em soja. A sequência para síntese dos marcadores foi obtida em SOYBASE (<http://soybase.agron.iastate.edu/>). Os fragmentos foram amplificados em um volume final de 10  $\mu\text{L}$  contendo: tampão de PCR 1X [Tris-HCl 2,5 mM (pH 8,3) e KCl 62,5 mM]; 2,5 mM de  $\text{MgCl}_2$ ; 125  $\mu\text{M}$  de cada dNTP; 0,2  $\mu\text{M}$  de cada oligonucleotídeo, uma unidade de *Taq* DNA polimerase e 20 ng de DNA genômico. As condições de amplificação foram: 7 min a  $95^{\circ}\text{C}$ , 29 ciclos de 1 min a  $94^{\circ}\text{C}$ , 1 min a  $50^{\circ}\text{C}$  e 1 min a  $72^{\circ}\text{C}$  e extensão de 7 min a  $72^{\circ}\text{C}$ . Os fragmentos foram separados por eletroforese em gel de poliacrilamida (29:1, acrilamida:bis-acrilamida) a 10%. As amostras foram coradas em solução de brometo de etídeo (0,5  $\mu\text{g/mL}$ ), fotografadas sob luz UV, com câmera Kodak CD 120 em um transluminador, com software Kodak Digital Scienc 1D (Eastman Kodak Company).

Todas as cultivares foram tratadas como linhagens puras. Uma pequena proporção de heterozigotos foi observada e o alelo mais comum foi considerado na análise. A partir do conhecimento prévio de que alguns dos marcadores estavam ligados, oito marcadores não ligados foram selecionados para a análise da estrutura populacional. O método descrito por PRITCHARD et al. (2000) e implementado no *software* Structure foi utilizado para testar a hipótese de 1 a 10 subpopulações ( $K$ ) (WANG et al., 2008), com mistura e frequências alélicas correlacionadas. Foram estabelecidos os parâmetros de 10 corridas para cada valor de  $K$ , com 10.000 interações em cada corrida, e 20.000 interações adicionais.

## Resultados e Discussão

Ao todo 192 cultivares foram genotipadas para 21 loci SSR em 13 cromossomos e o número de alelos variou entre 2 e 11, com uma média de 4,3 alelos por locus. Somente dois marcadores não revelaram polimorfismo. Na análise da estrutura populacional (Figura 1), para  $K=2$  foram obtidos dois grupos distintos. À medida que o valor de  $K$  aumentou, um novo grupo foi acrescentado aos demais. No entanto, para  $K=4$ , o acréscimo de novo grupo resultou na perda da distinção entre os grupos. Neste último caso, as cultivares apresentaram adesão parcial em vários grupos, com coeficientes de adesão semelhantes para a maioria das cultivares. Estas populações podem refletir gradações contínuas nas frequências alélicas nos grupos, conforme sugerido por ROSENBERG et al. (2002).

O número de marcadores utilizados na análise da estrutura populacional foi adequado. De acordo com PRITCHARD et al. (2000), o método implementado no *software* Structure é capaz de produzir resultados altamente acurados usando poucos loci. Os autores utilizaram sete loci SSR em estudo populacional de aves em extinção.

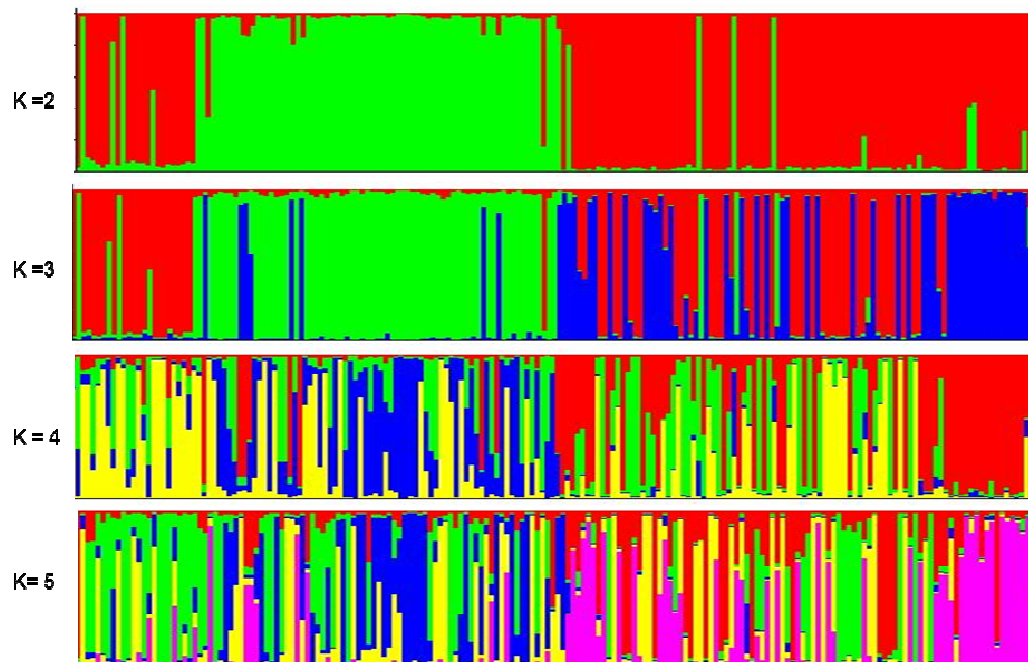
Os resultados confirmaram a base genética estreita da soja comercializada no Brasil. Além disso, devido ao método usualmente adotado pelas empresas privadas para obter ou desenvolver germoplasma, duas linhagens podem frequentemente possuir o mesmo *pedigree* (WANG et al., 2008). Portanto, é importante considerar a relação entre os membros das populações e considerar a estrutura e o parentesco.

**Tabela 1.** Distribuição de marcadores SSR para teor de proteína em soja.

| Grupo de ligação | Marcador  |
|------------------|---|
| A1               | <u>Satt 050</u> (46,45); <u>Satt 385</u> (64,74)                      |
| A2               | Satt 455 (129,86)   |
| B1               | <u>Sat 270</u> (21,99); <u>Satt 251</u> (36,48); Satt 509 (32,51)     |
| B2               | Satt 416 (56,96)  |
| C1               | Satt 180 (127,77); Satt 338 (123,79)                                  |
| C2               | Satt 202 (126,24); <u>Satt 277</u> (107,59); <u>Satt 357</u> (151,91) |
| D1a              | Satt 531 (40,87)  |
| H                | Satt 192 (44,04)  |
| I                | <u>Satt 239</u> (36,94); <u>Satt 571</u> (18,5)                       |
| J                | Satt 406 (38,19)  |
| M                | Satt 567 (33,47)  |
| N                | Sat_084 (36,86); Satt 159 (27,13)                                     |
| O                | Satt 478 (71,1)   |

Entre parênteses está a localização no mapa, obtida em SOYBASE (<http://soybase.ncgr.org/cgi-bin/ace/generic/search/soybase>) em 2/01/2010.

Os marcadores sublinhados foram selecionados para análise da estrutura populacional.



**Figura 1.** Estrutura populacional estimada com 192 cultivares de soja comercializadas no Brasil. Cada cultivar está representada por uma linha vertical distribuída em segmentos de  $K$ . Os valores de  $K$  correspondem ao número de subpopulações, representadas na figura por cores distintas.

## Conclusões

A população está estruturada e pode ser utilizada em estudos de mapeamento associativo.

## Referências

- CREGAN, P.B.; JARVIK, T.; BUSH, A.L.; SHOEMAKER, R.C.; LARK, K.G.; KAHLER, A.L.; KAYA, N.; VAN TOAL, T.T.; LOHNES, D.G.; CHUNG, J.; SPECHT, J.E. An integrated genetic linkage map of the soybean genome. **Crop Science**, v.39, p.1464 - 1491, 1999.
- HUNGRIA, M.; CAMPO, R.J.; MENDES, I.C. **A importância do processo de fixação biológica do nitrogênio para a cultura da soja: componente essencial para a competitividade do produto brasileiro**. Londrina: EMBRAPA/CNPQSO, 2007. 80 P. (Documentos 283).
- KEIM, P.; OLSON, T.C.; SHOEMAKER, R.C. A rapid protocol for isolating soybean DNA. **Soybean Genetics Newsletter**, v.15, p.150-152, 1988.
- NICOLÁS, M.F.; HUNGRIA, M.; ARIAS, C.A.A. Identification of quantitative trait loci controlling nodulation and shoot mass in progenies from two Brazilian soybean cultivars. **Field Crops Research**, v.95, p.355-366, 2006.
- PRITCHARD, J.K.; STEPHENS M.; DONNELLY, P. Inference of population structure using multilocus genotype data. **Genetics**, v.155, p.945-959, 2000.
- ROSENBERG, N.A.; PRITCHARD, J.L.; WEBER, J.L.; CANN, H.M.; KIDD, K.K.; ZHIVOTOVSKY, L.A.; FELDMAN, M.W. Genetic structure of human populations. **Science**, v.298, p.2381-2385, 2002.
- SANTOS, M.A.; NICOLÁS, M.F.; HUNGRIA, M. Identificação de QTL associados à simbiose entre *Bradyrhizobium japonicum*, *B. elkanii* e soja. **Pesquisa Agropecuária Brasileira**, v.41, p.67-75, 2006.
- SOYBASE 2010. **Soybase and the soybean breeder's toolbox. Knowledge an tools for soybean breeders and researchers**. USDA-Iowa State University, IA, USA. Disponível em: <http://soybase.agron.iastate.edu/>. Acesso em: 2 de jan. 2010.
- TANYA, P.; SRINIVES, P.; TOOJINDA, T.; NAKHON, P.; VANAVICHIT, A.; LEE, S.H.. Identification of SSR markers associated with N<sub>2</sub> fixation components in soybean [*Glycine max* (L) Merr.]. **The Korean Journal of Genetics**, v.27, p.351-359, 2005.
- WANG, J.; McCLEAN, P.E; LEE, R.; GOOS, R.J.; HELMS, T. Association mapping of iron deficiency chlorosis loci in soybean (*Glycine max* L. Merr.) advanced breeding lines. **Theoretical and Applied Genetics**, v.116, p.777-787, 2008